Combining multi-scale spatial data

Prepared for the

Workshop on Spatial Statistics
For Agricultural and Environmental Applications

by

Mark West
USDA/ARS/NPA

Combining spatial data from different sources is a common problem and poses a real challenge. (Gotway C.A. 2002) provides an overview of the most recent approaches and progress made towards combining incompatible spatial data. What is covered in this presentation is narrowly focused on available geostatistical approaches which are challenging enough. By illustrating a few examples, my aim is to introduce basic concepts and terminology used in geostatisitical analysis and expose the limitations of these methods.

Gotway C.A., Y. L. J. (2002). "Combining Incompatible Spatial Data." Journal of the American Statistical Association **97**(458): 632-648(17).

What are multi-scale data?

- multiple sources
- collected from the same region using different formats and scales
- each source (layer) may have one or more attributes (weed infestation, percent bare ground)
- sources may have different levels of accuracy and precision

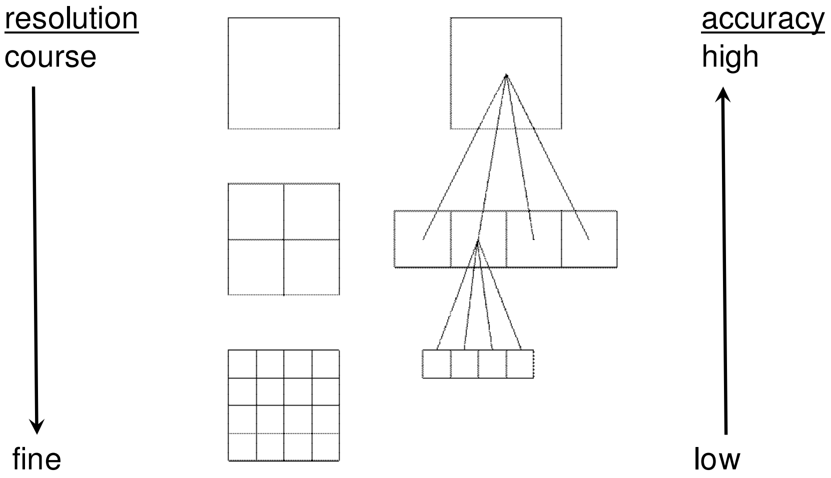
Multi-scale data may include different formats such as points, lines, polygons and grids. Edzer Pabesma (Pebesma 2004) has developed the **sp** and the **gstat** packages with R code for analyzing different types of spatial data.

(Zhu, Morgan et al. 2004) combine soil coring, penetrometer, and other topographic data to produce a fine map of depth-to-till for a Wisconson field. The data collected using these methods have different resolutions and accuracies. Soil coring provides accurate information on depth-to-till but because of its expense requires this information to be collected sparingly and hence results in a low resolution map of a field. Soil electroconductivity (EC) can provide information on depth-to-till and is easy to collect hence its resolution will be finer than the information collected from soil coring but has the problem of being less accurate with more error. The soil core and soil EC are multi-scale data.

Pebesma, E. J. (2004). "Multivariable geostatistics in S: the gstat package." Computers & Geosciences **30**(7): 683.

Zhu, J., C. L. S. Morgan, et al. (2004). "Combined mapping of soil properties using a multi-scale tree-structured spatial model." Geoderma **118**(3-4): 321.

Layers, Resolution & Accuracy



This slide was used to visually support the concept that resolution and accuracy aren't necessarily one in the same. Certainly we want to create maps that have fine resolution and are accurate. However, data that are easier to collect for providing finer maps are easier because the methodology used to collect them is quick, inexpensive and prone to error.

General Problem

- Inference at the level of one layer may be desired using information gathered from other levels
- Question may be "How do soil attributes measured at point locations relate to weed infestation measured on rectangular units?"

(Gotway 2002) gives several examples where data is on one scale but inference is desired at another. Individual level inference is wanted but because of privacy issues data is only available at some aggregate level. Data from Standard Metropolitan Statistical areas may be available but information at the county level may be needed.

Gotway, C. A. w. Young, L.J. (2002). "Combining Incompatible Spatial Data." Journal of the American Statistical Association **97**(458): 632-648(17).

Focus

- Use topographic and soil attributes to predict crop yield
- Predict at aggregate levels
- Mapping applications
- Use available geostatistical (kriging) methods
  - R programs
    - Packages (Edzer Pabesma, 2005)
      - **gstat**
      - **sp**

The focus of this presentation is to examine a few geostatistical methods (mainly kriging systems) that involve problems with combining misaligned spatial data. Definitions of terms related to the problems will also be given.

My intention is to provide information of available tools that can be used to krige data. These are available for free from the R Development Core Team (2005) . All of the kriging systems were fitted using the gstat package (Pebesma 2004) . There is a variety of example code for fitting similar systems in the gstat package. You need to download the package and once downloaded refer to the directory C:\Program Files\R\R-2.2.1\library\gstat\demo for example code. I found these scripts very helpful.

Pebesma, E. J. (2004). "Multivariable geostatistics in S: the gstat package." Computers & Geosciences **30**(7): 683.

R Development Core Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org

Example 1: Field scale study relating elevation to yield

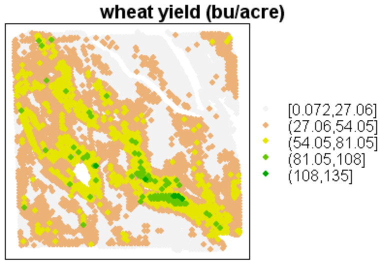
Models were fit that use elevation to predict spatial crop yield values. (Green & Erskine, 2004)

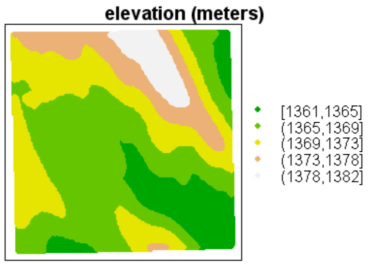
Can elevation help predict yield for large plots (blocks) in the field?

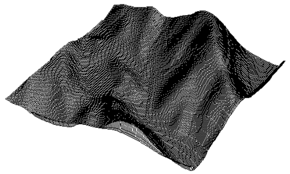
(Green 2004) addresses quantification of spatial variability of crop yield and soil water at farm scales using geostatistical and fractal analyses. His data are used in this example to demonstrate kriging methods for predicting wheat yield at the particular Nothern Colorado farm.

Timothy R. Green, R. H. Erskine. (2004). "Measurement, scaling, and topographic analyses of spatial crop yield and soil water content." Hydrological Processes **18**(8): 1447-1465.

Yield and elevation maps







Yield data (Green 2004) was collected from a field of roughly 800 square meters area using a combine mounted with a calibrated monitoring device to measure yield in bushels per acre. A GPS system was used to mark the 6701 points on the field where yield was recorded. Each yield value represents around 10 square meters of area. Because of various factors (e.g. the combine will not move at a regular speed) the actual area represented at each point will vary and yield values can be expected to be quite noisy.

Elevation data was collected over the same area using an all terrain vehicle. These data were interpolated to a regular grid of 5 meter spacing. These interpolated values of elevation should be very accurate representing small deviations from the true elevation (~ 0.05 m).

Yield and elevation maps were generated using the **sp** (Pebesma 2005) package in R.. See 'wheat yield and elevation plots.R' in the wheat folder on the ftp site.

Green, T. R., Erskine, Robert H. (2004). "Measurement, scaling, and topographic analyses of spatial crop yield and soil water content." Hydrological Processes **18**(8): 1447-1465.

Pebesma, E. J., Bivand, Roger S. (2005). S Classes and Methods for Spatial Data:the sp Package.

Example 2: Precision agriculture

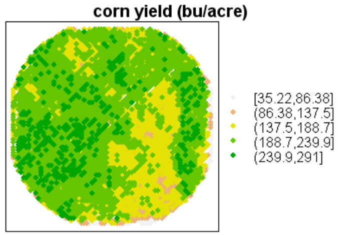
Soil electroconductivity (EC) mapping to explain yield variability for a center pivot cropping system in Northern Colorado.

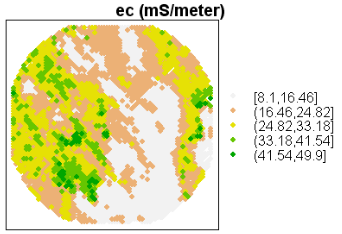
Can EC help predict yield for large blocks in the center pivot system?

The Water Management Research Unit in Fort Collins develops irrigation, agricultural chemical, and other management practices that protect water quality for all Americans while improving the husbandry of natural resources and the irrigator's economic viability. Research covers precision farming with center pivot sprinklers, remote sensing, and weed management for reduced applications of chemicals.

The data given here was collected in 1999 for relating various soil properties with soil electroconductivity (EC). Yield data was collected in 1999. Each point roughly represents around 11 to 12 sq. meters (the swath length is 20 ft.; the distance between points is around 6 ft).

Yield and electroconductivity (EC) maps





This data is similar the wheat and elevation data. The data given here was collected in 1999 for relating various soil properties with soil electroconductivity (EC). Yield values represent around 11 to 12 sq. meters. The EC data are measured can be expected to be much noisier than the elevation data.

Yield and EC maps were generated using the **sp** (Pebesma 2005) package in R.. See 'corn yield and EC plots.R' in the corn folder on the ftp site.

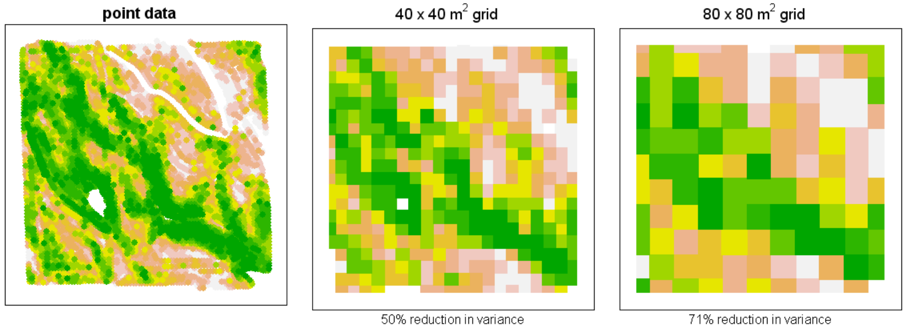
Pebesma, E. J., Bivand, Roger S. (2005). S Classes and Methods for Spatial Data:the sp Package.

Methods

- Aggregation
- Change of Support (COSP)
- Kriging
  - Point Kriging
  - Cokriging models
  - Block kriging
- Spatial joins

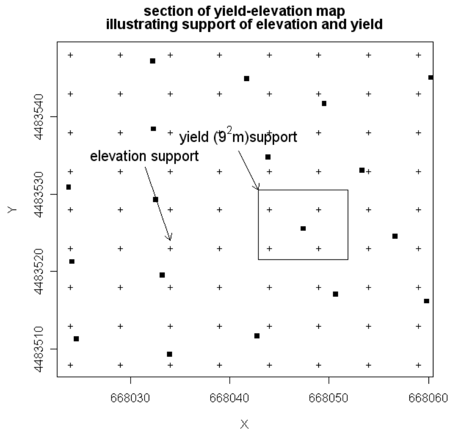
Methods used to combine multi-scale spatial data include aggregation, various kriging methods and those that involve what is referred to as a change of support (block kriging). We may be interested in changing from a point system to a system of blocks, from a system of blocks to a system of points., or from a system of blocks to another system of blocks.

Aggregation –
averaging over point values to form areal units



It is a well-known fact of statistics that averaging reduces variance. The apparent spatial variation also changes with aggregation.

Support of data



This slide is provided to point out data can be recorded at points but may have areal support. The yield data in both examples are geo-referenced at points but because the grain collected by the combine is collected over a region it represents yield over some small area (around 10 square meters for the wheat yield and around 11 square meters for the corn yield).

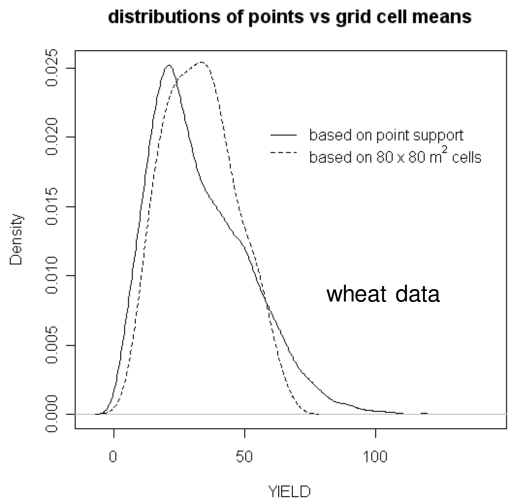
Why Aggregation?

- **Prediction:** prediction is wanted at larger scales
- **Different support:** aggregation transforms a variable from point support to areal support.
- **Smoothing:** aggregation smoothes out noise to detect trends

It isn't uncommon where data is collected at one scale and inference is desired at another. Making inferences on block averages whose support is different from those of the data is called a change of support problem. (Isaaks and Srivastava 1989) give an example of a mining operation where data are collected at points but mining operations involve only large blocks of material extracted from the mine. Having only point data on hand the problem here is to estimate the distribution of the average tonnage of ore contained in blocks. To estimate this sampled point data need to be aggregated to the size of blocks and the distribution of values associated with blocks may then be used to base decisions. By aggregation we mean obtaining a weighted average. To estimate the average tonnage of ore Z_B for a block B we need to come up with an estimate based on sampled values Z_i in the neighborhood of block. The estimator $\hat{Z}_B = \sum_i \lambda_i Z_i$ is derived by choosing weights λ_i that account for the spatial variation in the Z_i and the estimated spatial variation occurring on the block scale. Spatial variation for the change of support problem is modeled through a variogram $\gamma(h)$ of the Z_i.

Isaaks, E. H. and R. M. Srivastava (1989). Applied geostatistics. New York, Oxford University Press.

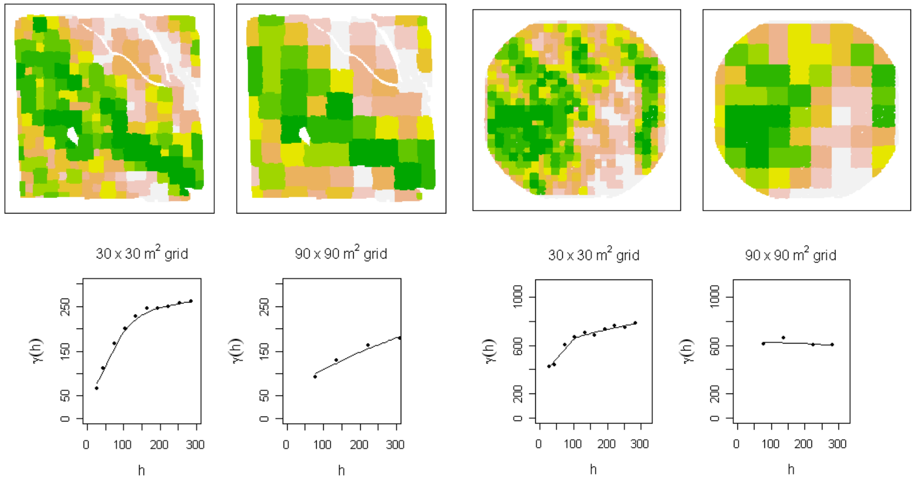
Support effect



The support effect is the change of distribution of statistics that results when data are aggregated. Quoting from (Gotway 2002), 'Changing the support of a variable (typically by averaging or aggregating) creates a new variable. This new variable is related to the original one, but has different statistical and spatial properties,'

Gotway, C. A. w. Y., L.J.) (2002). "Combining Incompatible Spatial Data." Journal of the American Statistical Association **97**(458): 632-648(17).

Scale Problem



The effect of the change of support realized by increasing the areas for which aggregation is performed is called the scale effect. The differences in the statistical properties of the variograms are obvious with scale changes for both examples.

Problems of Aggregation

- Change of support problem (COSP)
  - How can spatial variation at the point support scale be used to estimate spatial variation at an aggregate scale?
  - COSP modeled through variogram
  - Similar to using population variance to form inferences using sample means.

$$\sigma_y^2 \rightarrow \sigma_{\bar{y}}^2 = \sigma_y^2 / n$$

Making inferences on block averages whose support is different from those of the data is called a change of support problem. (Isaaks and Srivastava 1989) give an example of a mining operation where data are collected at points but mining operations involve only large blocks of material extracted from the mine. Having only sampled point data available a big problem is to estimate the distribution of the average tonnage of ore contained in blocks.

Isaaks, E. H. and R. M. Srivastava (1989). Applied geostatistics. New York, Oxford University Press.

Aggregation Methods

- **Generic** : weighted average of values Z_i for estimating average for an area B of size |B|
  $\hat{Z}_B = \sum_i \lambda_{B_i} Z_i, \ \sum_i \lambda_{B_i} = 1$
- **Arithmetic means:** simple averages ($\lambda_{B_i} = 1/n$, n= sample size) ignore spatial structure
- **Kriging:** averages use weights λ_{B_i} derived from spatial structure $\gamma(h)$ - variogram

By aggregation we mean obtaining a weighted average. To estimate the average Z_B of a variable Z for a block B we need to come up with a weighted average $\hat{Z}_B = \sum_i \lambda_{B_i} Z_i$ based on sampled values Z_i in the block neighborhood. Later more detail will be given when the method of block kriging is described.

Some Notation

- S = point where an observation is made
- Z(S) = value of observation at S
- $\delta(\mathbf{s})$ = error from mean value at S
- μ = mean value at for any S
- $\mu(\mathbf{s})$= mean value that depends on location S and/or predictors at S

Notation added to clarify expressions to follow.

Ordinary and Universal Kriging

Ordinary Kriging

–Model: $Z(\mathbf{s}) = \mu + \delta(\mathbf{s})$

- Universal Kriging

–Model: $Z(\mathbf{s}) = \mu(\mathbf{s}) + \delta(\mathbf{s})$

- Predictor: $Z(S)=\sum_i \lambda_i \cdot Z(S_i)$
  - **λ_i weight of i^{th} value, derived from variogram of $\delta(S)$ and/or predictors**

Kriging is a method of spatial prediction. The predictors are in the form of a weighted average $\hat{Z} = \sum_i \lambda_i \cdot Z_i$. The differences in these two kriging methods are their underlying models.

For ordinary kriging, the underlying model for the Z is a constant mean plus error where errors are spatially autocorrelated. The spatial autocorrelation of errors doesn't depend on location. The λ_i are derived using the model assumptions to give the minimum mean-squared prediction error. For ordinary kriging, the λ_i are a function of the variogram $\gamma(h)$ that describes the autocorrelation of errors.

For universal kriging, the underlying model for the Z is a mean that depends on location and/or other predictor variables plus error where the errors are spatially autocorrelated. Again, the spatial autocorrelation of errors doesn't depend on location. For universal kriging, the λ_i are a function of the variogram $\gamma(h)$ that describes the autocorrelation of errors and the predictors that are modeling the mean.

Cokriging

- Simultaneously krige two or more variables

$Z(s)=\sum_i \lambda_i \cdot Z(s_i) + \sum_j \omega_j \cdot X(u_j)$

$Z(s_i)$ yield at locations s_i

$X(u_j)$ EC at locations u_j

- Not only requires fitting of variograms for each variable but also requires fitting of the cross-variogram for each pair of variables

Cokriging is a method originating from the need for predicting a primary variable Z that is undersampled (because it may be expensive to sample) but another secondary variable X is available that is related to Z and more heavily sampled (because X it is less expensive/difficult to sample). Both X(S) and Z(S) are fitted to a model simultaneously. This is a form of multivariate prediction modeling. The estimator for an unknown Z is of the form $\hat{Z} = \sum_i \lambda_i \cdot Z_i + \sum_j \omega_j \cdot X_j$. The usefulness of the secondary variable for predicting the primary variable is enhanced when the primary is undersampled. See (Isaaks and Srivastava 1989) for a more complete description.

Isaaks, E. H. and R. M. Srivastava (1989). Applied geostatistics. New York, Oxford University Press.

Block Kriging

- estimate the mean value of an attribute for a local area **B** using points in the neighborhood of **B**
- used with either ordinary, universal or cokriging
- variogram is adjusted to handle the scale effect
- estimator: $\mathbf{\hat{Z}(B)=\sum_i \lambda_{B_i} \cdot Z(s_i)} \approx \frac{1}{|B|}\int_B Z(s)ds$

Block kriging is an aggregation method for estimating or predicting an average value $\hat{Z}_B$ over an area B. The estimator $\hat{Z}_B = \sum_i \lambda_{B_i} Z_i$ is derived by choosing weights λ_{B_i} that account for the spatial autocorrelation in the Z_i and the estimated spatial autocorrelation occurring on the block scale. Therefore we need to know how the autocorrelation among units on the point scale changes to autocorrelation among units on the block scale. Spatial variation for the change of support problem is modeled through a variogram $\gamma(h)$ of the Z_i .(Cressie 1993) describes the needed calculations to modify the point support variogram $\gamma(h)$ to the block support variogram $\gamma(B)$ (pages 124-125,) for block kriging (aggregation over an area B).

Cressie, N. A. C. (1993). Statistics for spatial data. New York, J. Wiley.

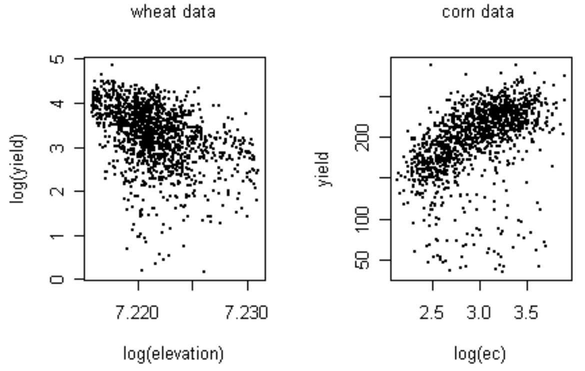
Spatial Join of datasets

- Combine two or more datasets with different attributes measured at different locations by translating them to same location.
- Problems
  - Trans-locating errors in variables problem
  - Ad Hoc approach, descriptive purposes
- Benefits
  - plotting techniques reveal relationships and needed transformations for other more legitimate methods
  - attributes can be studied together

Data measured at different locations can be joined many different ways. Consider two geostatistical datasets A and B each with different attributes. One approach would be to conduct a search of points in dataset B for each data point in dataset A. The attribute values corresponding to the points in B nearest in distance to those in A are joined with those of A. Another approach would be to lay a grid over the intersection of the areas from which datasets are formed. For each point in the grid, a search is conducted to find the points in A and in B that are closest and these two are joined. Yet another way would be to spatially interpolate all the points in B to those in A. Many Geographic Information Systems provide software for joining misaligned data but the capabilities of the software is limited to descriptive purposes.

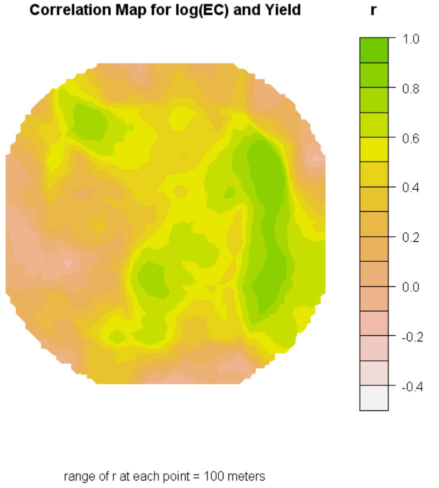
See **joindata.R** for an R program that quickly joins two datasets.

Scatterplots of spatially joined data



When kriging methods involve predictors, they are linear functions of the predictors. Spatial joins were used to construct scatterplots to study the relationships between yield and predictors to find transformations that ensure linearity. In both cases, suitable transformations were found.

A spatial join may be used to see if relationships are stationary



After creating a spatial join, Pearson's r was calculated as a moving window statistic to consider how the relationship between EC and YIELD may change throughout the region

Using spatially joined data from the center pivot example, Pearson's correlation coefficient was used as a 'moving-window' statistic. A 12x12 m grid was overlaid on the center-pivot area. For each point on the resulting grid, all points in the spatially joined dataset within a 100 m radius were selected and the correlation coefficient was computed, and then mapped.

(Carroll and Oliver 2005) give details of this technique in their study of EC and soil properties.

Carroll, Z. L. and M. A. Oliver (2005). "Exploring the spatial relations between soil physical properties and apparent electrical conductivity." Geoderma **128**(3-4): 354.

Attribute characteristics

- Wheat Yield / Elevation
  - Response: **Yield**
  - Predictor: **Elevation, measured with little error**
  - location: measured at different locations
- Corn Yield / EC
  - Response: **Yield**
  - Predictor: **EC, measured** with a lot of error
  - location: measured at different locations

Before looking at the specific problems for using elevation or EC to help predict yield using kriging methods, this slide is given to motivate why different approaches for predicting yield are taken

For the wheat yield data, elevation has so little error associated with it that it is practical to treat these values as being static. There will be little error incurred by interpolating elevation values to points where wheat yield is observed. Doing this we act as if both yield and elevation are measured at the same points in the field. I view this problem as a univariate regression problem where the predictor, elevation, is known for any point in the field

For the corn yield data, soil EC has a lot of error associated with it to begin with. Interpolating EC value to points where corn yield is observed will add more error. For practical as well as illustrative reasons corn yield and soil EC joined to the same points for analysis. I view this problem differently in that it lends itself to a cokriging application.

Suggested approach using elevation to help predict yield

- Universal Kriging
  - interpolate values of the elevation to locations where yield is recorded
  - use elevation as a predictor

Using the example datasets, two approaches are considered for incorporating the predictor variables with kriging methods.

For the wheat yield example, universal kriging will seems to be a reasonable approach for predicting yield using the model $Z(S) = \mu(S) + \beta \cdot X(S) + \delta(S)$ where $Z(S) = \log(\text{Yield})$ at location S, $\mu(S)$ is the mean value of log(Yield) at location S, $X(S) =$ interpolated value of log(Yield) at location S and $\delta(S)$ is the error at location S. Although elevation is not observed at the same locations as yield, interpolating elevation to those points where yield is measured should incur little error.

For the corn yield example, universal cokriging will be used on the basis that a reasonable model for predicting yield is $\underline{Z}(S) = \underline{\mu}(S) + \underline{\delta}(S)$ where $\underline{Z}(S)$ represents the bivariate values of both yield and EC at location S, $\underline{\mu}(S)$ represents the mean of the bivariate values at location S and $\underline{\delta}(S)$ represents the bivariate errors at location S.

Predicting corn yield Can EC help prediction?

- Hold out thirty 40 x 40 m^2 grid cells for comparison of prediction methods
- Use remaining data to fit prediction models; **ordinary kriging**, **universal kriging**, and **universal cokriging**
- Block krige yield to 40 x 40 m^2 grid using each method & obtain standard errors

Steps used to compare kriging methods for the corn yield data are similar to that of the wheat yield only yield data and EC data weren't spatially joined.

The usefulness of universal cokriging for predicting blocks of unsampled plots was tested by holding out a set of thirty randomly selected 40 x 40 m^2 blocks, predicting their average values and comparing them back to the actual means for those blocks. For comparison purposes, ordinary kriging and universal kriging using locations were included. Using each kriging method, yield was block kriged to predict average values for the 40 x 40 m^2 held out blocks. The abilities of these methods for prediction were evaluated by comparing r^2. Standard errors of the estimates were also compared.

The spatial autocorrelation structure of the errors for the model was fitted to Gausian variogram models. For universal kriging, residuals were obtained by fitting a trend surface of yield over the field. The residuals were then used to obtain an empirical variogram. The empirical variogram was fitted to a Gaussian variogram model by least squares.

Kriging approach using EC to help predict yield values on blocks

- Cokriging
  - Simultaneously krige both yield and EC
  - Fit linear model of coregionalization (LMC)
    - a method for fitting variograms for yield and EC and the cross-covariogram of EC and yield

$$\mathbf{Z(s_0)=\sum_i \lambda_i \cdot Z(s_i) + \sum_j \omega_j \cdot X(u_j)}$$

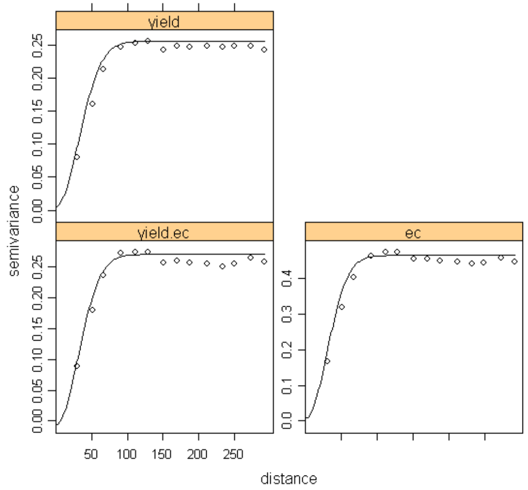
$Z(s_i)$ yield at locations s_i

$X(u_j)$ predictor at locations u_j

I chose to explore the ability of cokriging yield and EC because I felt this to be a reasonable application to cokriging. I felt this to be a more reasonable application than cokriging yield and elevation since the error in elevation was expected to be very small. A linear model of coregionalization was used to fit the variograms and cross-variogram for yield and EC. The method is described in (Isaaks and Srivastava 1989) and as a word of warning can be quite an undertaking.

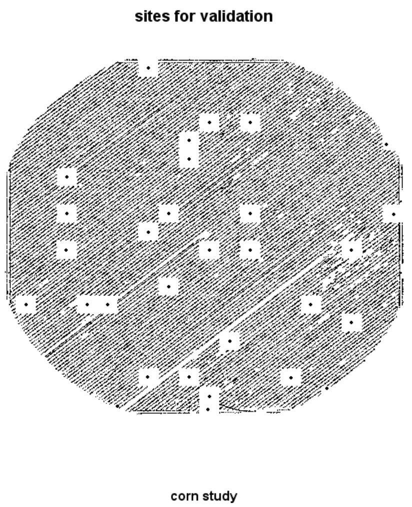
Isaaks, E. H. and R. M. Srivastava (1989). Applied geostatistics. New York, Oxford University Press.

Fitted cross-variogram for cokriging



Cokriging requires fitting of the cross-variograms(semivariograms) as well as the variograms.

Validation Sites



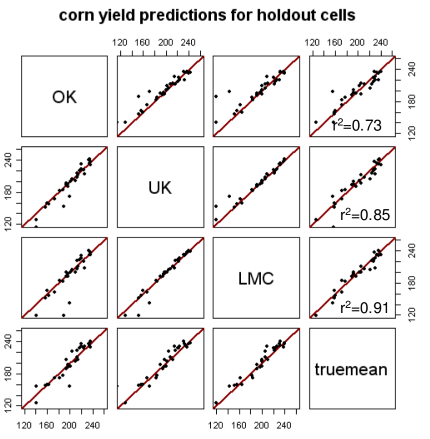
Random samples of thirty 40 x 40 m2 blocks of data were held out in order to compare methods. The plot shows the locations of these plots.

Compare methods

- OK: Ordinary kriging
- LMC: cokriging with linear model of coregionalization
- UK: Universal kriging
- True Mean: Observed average yield

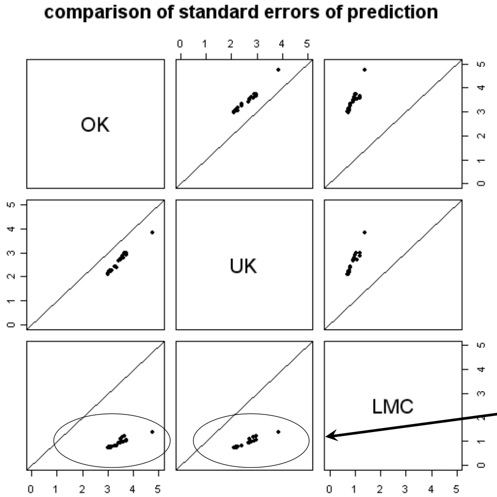
To follow methods of prediction, abbreviations are made. Kriging methods compared for predicting the thirty held out plots in the field are; **OK** – ordinary kriging; **UK** – universal kriging using locations as predictors; **LMC** – universal cokriging of yield and EC using a method of fitting a cross-covariogram called linear model of coregionalization; **truemean** – is used to denote the actual sample average observed for the held out plots. See (Isaaks and Srivastava 1989) for details of the linear model of coregionalization.

Compare predictions



The scatterplot matrix is used to make comparisons among the predictions for the means of the holdout plots. Each of the methods compared are weighted averages of values at points in the neighborhood of the plot being estimated. All of the estimates agree well with the true sample averages for the plots being estimated. r^2 is calculated by squaring the correlation coefficient computed between the true sample average for each of the thirty holdout plots and the corresponding prediction values obtained by each method: OK – ordinary kriging; UK – universal kriging using locations as predictors; LMC – universal cokriging of yield and EC using a method of fitting a cross-covariogram called linear model of coregionalization .See (Isaaks and Srivastava 1989) for further details. Based on the r^2, there appears to be some increased ability of prediction over ordinary kriging using the universal and cokriging methods. However, this may be due to the search region of points used for making the predictions.

Compare standard errors of methods used



Standard errors for LMC smaller than for OK and UK

Ordering the methods from largest to smallest on the basis of size of standard errors, ordinary kriging is showing the largest standard errors, then the next largest is for universal kriging and the smallest is for universal cokriging. Intuition would lead you to think this would be the order relation for these standard errors. The more information used would lead to more precise estimates. However, I am wary of these estimates because of the personal choices I made in the fitting of the variograms used for developing these universal kriging predictors. As a check I computed approximate 95% confidence intervals by calculating estimate +/- 2 standard errors, and found the proportion of intervals that cover the true means to be 3% with the LMC. An adjustment to the variogram estimates brought this coverage up to 100% with the standard errors still smaller than those for the ordinary kriging method.

Summary

- kriging is spatial prediction tool that uses weighted averages
  - weights depend on autocorrelation structure
  - explanatory variables may adjust the weights for a more accurate prediction - cokriging or universal kriging
- cokriging and universal kriging are ways to incorporate multi-scale data
- aggregation methods compared here give similar predictions but some accuracy and precision may improve with predictors
- block kriging is a useful scaling tool
- joining misaligned spatial data may be useful as a exploratory/descriptive tool
- examples given involve heavily sampled spatial regions– with less heavily sampled data a predictor may have bigger impact
- focus was on geostatistical methods, newer Bayesian methods namely tree-structured hierachical models may be more effective (Zhu 2004, 2005)

To summarize, geostatistical methods covered here mostly revolve around mapping applications. Although difficult, different sources of data can be combined to improve mapping accuracy and precision.

Thank You